0 3/36003

PCT

(51) International Patent Classification 4 : C07H 15/12, C12P 21/00, 19/34 C12N 15/00, 7/00, 1/20 C12Q 1/68, C07K 13/00	A1	(11) International Publication Number: **WO 89/09777**
		(43) International Publication Date: 19 October 1989 (19.10.89)

(21) International Application Number: PCT/US89/01473

(22) International Filing Date: 7 April 1989 (07.04.89)

(30) Priority data:
179,587   8 April 1988 (08.04.88)   US
249,584   26 September 1988 (26.09.88)   US

(71) Applicant: ARCH DEVELOPMENT CORPORATION [US/US]; 1115-25 East 58th Street, Chicago, IL 60637 (US).

(72) Inventor: SUKHATME, Vikas, P. ; 1511 East 56th Street, Chicago, Il 60637 (US).

(74) Agent: MEINERT, M., C.; Marshall, O'Toole, Gerstein, Murray & Bicknell, Two First National Plaza, Ste. 2100, Chicago, IL 60603 (US).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS AND MATERIALS RELATING TO DNA BINDING PROTEINS

(57) Abstract

DNA sequences associated with regulation of early stages of cell growth are described. Illustratively provided are human and mouse origin DNA sequence encoding early growth regulatory ("Egr") proteins which include "zinc finger" regions of the type involved in DNA binding. Immunological methods and materials for detection of Egr proteins and hybridization methods and materials for detection and quantification of Egr protein related nucleic acids are also disclosed.

- 1 -

"METHODS AND MATERIALS RELATING TO
DNA BINDING PROTEINS"

5

CROSS-REFERENCE TO RELATED APPLICATIONS

10   This is a continuation-in-part application of
co-pending U.S. Patent Application Serial No. 179,587,
filed April 8, 1988.

15   BACKGROUND OF THE INVENTION

        The present invention relates generally to DNA
20 binding regulatory proteins and more particularly to DNA
sequences encoding early growth regulatory proteins
possessing histidine-cysteine "zinc finger" DNA binding
25 domains, to the polypeptide products of recombinant
expression of these DNA sequences, to peptides and poly-
peptides whose sequences are based on amino acid
sequences deduced from these DNA sequences, to anti-
30 bodies specific for such proteins and peptides, and to
procedures for detection and quantification of such pro-
teins and nucleic acids related thereto.
35   Among the most significant aspects of mam-
malian cell physiology yet to be elucidated is the
precise manner in which growth factors (e.g., hormones,
neurotransmitters and various developmental and
differentiation factors) operate to effect the regula-
tion of cell growth. The interaction of certain growth
factors with surface receptors of resting cells appears
to rapidly induce a cascade of biochemical events
thought to result in nuclear activation of specific
growth related genes, followed by ordered expression of
other genes. Analysis of sequential activation and
expression of genes during the transition from a resting
state ("G_0") to the initial growing state ("G_1") has
been the subject of substantial research. See, gener-

- 2 -

ally, Lau et al., <u>Proc. Nat'l. Acad. Sci. (USA)</u>, <u>84</u>,
1182-1186 (1987). Much of this research has involved
analysis of the expression of known genes encoding sus-
pected regulatory proteins (such as the proto-oncogenes,
c-fos and c-myc) following mitogen stimulation. An
alternative approach has involved attempts to identify
genes activated by mitogenic stimuli through differen-
tial screening of cDNA libraries prepared from resting
cells following exposure to serum and specific growth
factors. See, e.g., Lau et al., <u>EMBO Journal</u>, <u>4</u>,
3145-3151 (1985). See also, Cochran et al., <u>Cell</u>, <u>33</u>,
939-947 (1983), relating to the cloning of gene
sequences apparently regulated by platelet derived
growth factor.

     Of interest to the background of the invention
is the continuously expanding body of knowledge regard-
ing structural components involved in the binding of
regulatory proteins to DNA. Illustratively, the so-
called receptor proteins are believed to bind to DNA by
means of zinc ion stabilized secondary structural
fingers premised on folding of continuous amino acid
sequences showing high degrees of conservation of
cysteines and histidines and hydrophobic residues. See,
e.g., Gehring, <u>TIBS</u>, <u>12</u>, 399-402 (1987). For example, a
"zinc finger" domain or motif, present in <u>Xenopus</u>
transcription factor IIIA (TF IIIA), as well as the
Drosophila Kruppel gene product and various yeast
proteins, involves "repeats" of about 30 amino acid
residues wherein pairs of cysteine and histidine
residues are coordinated around a central zinc ion and
are thought to form finger-like structures which make
contact with DNA. The histidine-cysteine (or "CC-HH")
zinc finger motif, as opposed to a cysteine-cysteine
("CC-CC") motif of steroid receptors, is reducible to a
consensus sequence represented as $C-X_{2-4}-C-X_3-F-X_5-L-X_2-H-X_3-H$ wherein C represents cysteine, H represents

- 3 -

histidine, F represents phenylalanine, L represents
leucine and X represents any amino acid. [See, Klug et
al. TIBS, 12, 464-469 (1987); Blumberg et al., Nature,
328, 443-445 (1987); and Schuh et al., Cell, 47,
1025-1032 (1986).]

Of particular interest to the background of
the invention is the recent report of Chowdhury et al.,
Cell, 48, 771-778 (1987), relating to an asserted
"family" of genes encoding proteins having histidine-
cysteine finger structures. These genes, designated
"mkr1" and "mkr2", appear to be the first such isolated
from mammalian tissue and are not correlated to any
early growth regulatory events.

There continues to exist a need in the art for
information concerning the primary structural conforma-
tion of early growth regulatory proteins, especially DNA
binding proteins, such as might be provided by knowledge
of human and other mammalian DNA sequences encoding the
same. Availability of such DNA sequences would make
possible the application of recombinant methods to the
large scale production of the proteins in procaryotic
and eukaryotic host cells, as well as DNA-DNA and
DNA-RNA hybridization procedures for the detection,
quantification and/or isolation of nucleic acids asso-
ciated with these and related proteins. Possession of
such DNA-binding proteins and/or knowledge of the amino
acid sequences of the same would allow, in turn, the
development of monoclonal and polyclonal antibodies
thereto (including antibodies to protein fragments or
synthetic peptides modeled thereon) for use in immuno-
logical methods for the detection and quantification of
early growth regulatory proteins in fluid and tissue
samples as well as for tissue specific delivery of sub-
stances such as labels and therapeutic agents to cells
expressing the proteins. In addition, DNA probes based
on the DNA sequences for these mammalian early growth

- 4 -

regulatory proteins may be of use in detecting gene
markers used for the diagnosis of those clinical
5   disorders which are linked to the marker genes.

BRIEF SUMMARY OF THE INVENTION

10   The present invention provides novel purified
and isolated DNA sequences encoding mammalian early
growth regulatory ("Egr") proteins which comprise one or
15   more histidine-cysteine zinc finger amino acid sequences
putatively providing DNA binding (and hence DNA replica-
tion or transcription regulatory) capacity.   In
20   presently preferred forms, novel DNA sequences of the
invention comprise genomic and cDNA sequences encoding
human and mouse early growth regulatory proteins.
Alternate DNA forms, such as "manufactured" DNA,
25   prepared by partial or total chemical synthesis from
nucleotides, are also within the contemplation of the
invention.
30   Operative association of Egr-encoding DNA
sequences provided by the invention with homologous or
heterologous species expression control DNA sequences,
35   such as promoters, operators, regulators and the like,
allows for in vivo and in vitro transcription to form
messenger RNA which, in turn, is susceptible to transla-
tion to provide Egr proteins in large quantities.   In
one presently preferred DNA expression system practiced
according to the invention, Egr-encoding DNA is opera-
tively associated with a bacteriophage T3 or T7 RNA
promoter DNA sequence allowing for in vitro transcrip-
tion and translation in a cell free system.   Incor-
poration of novel DNA sequences of the invention into
procaryotic and eucaryotic host cells by standard trans-
formation and transfection processes involving suitable
viral and circular DNA plasmid vectors is also within
the contemplation of the invention and is expected to

- 5 -

provide useful proteins in quantities heretofore
unavailable from natural sources. Illustratively, frag-
5   ments of DNA encoding Egr protein of the invention have
been incorporated in plasmid vectors resulting in
expression by transformed E.coli hosts of fusion pro-
teins sharing immunological characteristics of Egr
10   protein. Use of mammalian host cells is expected to
provide for such post-translational modifications (e.g.,
truncation, glycosylation, and tyrosine, serine or
15   threonine phosphorylation) as may be needed to confer
optimal biological activity on recombinant expression
products of the invention.

20   Also provided by the present invention are
novel, presumptively mitogen sensitive, DNA sequences
involved in regulation of the transcription of Egr-
encoding DNA, which sequences are expected to have
25   utility in the efficient recombinant expression of Egr
proteins as well as proteins encoded by other structural
genes. In addition, the DNA sequences may be used as
30   probes to detect the presence or absence of gene markers
used for the diagnosis of clinical disorders linked to
those gene markers.

35   Novel polypeptide products of the invention
include polypeptides having the primary structural con-
formation (i.e., amino acid sequence) of Egr proteins or
fragments thereof, as well as synthetic peptides, an
analogs thereof, assembled to be partially or wholly
duplicative of amino acid sequences extant in Egr
proteins. Proteins, protein fragments, and synthetic
peptides of the invention are expected to have
therapeutic, diagnostic, and prognostic uses and also to
provide the basis for preparation of monoclonal and
polyclonal antibodies specifically immunoreactive with
Egr proteins, as well as to provide the basis for the
production of drugs for use as competitive inhibitors or
potentiators of Egr. Preferred protein fragments and

- 6 -

synthetic peptides of the invention include those
duplicating regions of Egr proteins which are not
5   involved in DNA binding functions (i.e., regions other
than the zinc fingers). Most preferred are peptides
which share at least one continuous or discontinuous
antigenic epitope with naturally occurring Egr pro-
10   teins.

Antibodies of the invention preferably bind
with high immunospecificity to Egr proteins, fragments,
15   and peptides, preferably recognizing epitopes which are
not common to other proteins, especially other DNA bind-
ing proteins.

20   Also provided by the present invention are
novel procedures for the detection and/or quantification
of Egr proteins and nucleic acids (e.g., DNA and mRNA)
specifically associated therewith. Illustratively, anti-
25   bodies of the invention may be employed in known immuno-
logical procedures for quantitative detection of Egr
proteins in fluid and tissue samples. Similarly, DNA
30   sequences of the invention (particularly those having
limited homology to other DNAs encoding DNA binding
proteins) may be suitably labelled and employed for the
35   quantitative detection of mRNA encoding the proteins.
Information concerning levels of Egr mRNA may provide
valuable insights into growth characteristics of
cells.

Among the multiple aspects of the present
invention, therefore, is the provision of (a) novel
purified and isolated Egr-encoding DNA sequences set out
in Figures 1A, 3, and 4 as well as (b) Egr-encoding DNA
sequences which hybridize thereto under hybridization
conditions of the stringency equal to or greater than
the conditions described herein and employed in the
initial isolation of DNAs of the invention, and (c)
synthetic or partially synthetic DNA sequences encoding
the same, or allelic variant, or analog Egr polypeptides

which employ, at least in part, degenerate codons.
Correspondingly provided are viral or circular plasmid
5   DNA vectors incorporating such DNA sequences and pro-
caryotic and eucaryotic host cells transformed or trans-
fected with such DNA sequences and vectors as well as
novel methods for the recombinant production of Egr
10 proteins through cultured growth of such hosts and iso-
lation from the hosts or their culture media.

Preferred polypeptide products of the inven-
15 tion include those wholly or partially duplicating the
deduced sequence of the amino acid residues set out in
Figures 1A and 3 (i.e., mouse "Egr-1" and human
20 "EGR2"). Other preferred polypeptides include fusion
proteins such as cro-β-galactosidase/Egr-1 and bovine
growth hormone/Egr-1.

Presently preferred antibodies of the inven-
25 tion include those raised against synthetic peptides
partially duplicating deduced Egr amino acid sequences
of Figures 1A and 3 (e.g., the synthetic peptides H-L-R-
30 Q-K-D-K-K-A-D-K-S-C, the first 12 amino acid residues of
which duplicate mouse Egr-1 residues 416-427 with the
last cysteine added for coupling to KLH; and C-G-R-K-F-
35 A-R-S-D-E-R-K-R-H-T-K-I duplicating mouse Egr-1 residues
399-415). The antisera against the first peptide is
designated VPS10 and comprises a preferred antibody of
the invention.

As employed herein, the term "early growth
regulatory protein" shall mean and include a mammalian
DNA binding protein encoded by DNA whose transcription
temporally corresponds to cellular events attending the
G_0/G_1 growth phase transition. As employed herein,
"histidine-cysteine zinc finger amino acid sequence"
shall mean and include the following sequence of amino
acids $C-X_{2-4}-C-X_3-F-X_5-L-X_2-H-X_3-H$ wherein C represents
cysteine, H represents histidine, F represents phenyl-
alanine, L represents lysine, and X represents an amino
acid.

- 8 -

Other aspects and advantages of the present
invention will be apparent upon consideration of the
following detailed description thereof which includes
numerous illustrative examples of the practice of the
invention, reference being made to the drawing wherein:

Figure 1A provides a 3086 base nucleotide
sequence for a mouse Egr-1 DNA clone as well as a
deduced sequence of 533 amino acid residues for the
protein; Figure 1B provides a partial restriction map of
Egr-1 DNA clones together with information concerning
the position of the protein coding sequence and the
locus of amino acids providing for histidine-cysteine
zinc fingers;

Figure 2 provides an amino acid sequence
alignment of the DNA binding domain of mouse Egr-1 in
comparison with a zinc finger consensus sequence, with
the Drosophila Kruppel sequence and with the "finger 2"
sequence of Xenopus TFIIIA protein;

Figure 3 provides a 2820 base nucleotide
sequence for a human EGR2 cDNA clone as well as a
deduced sequence of 456 amino acids for the protein;

Figure 4 provides a 1200 base nucleotide
sequence of a mouse Egr-1 genomic clone, specifically
illustrating the 5' non-transcribed regulatory region
thereof comprising bases -935 through +1; and

Figure 5 provides a restriction map and
organization of the mouse Egr-1 genomic clone mgEgr-1.1
and a comparison to mouse Egr-1 cDNA.

DETAILED DESCRIPTION

The following examples illustrate practice of
the invention. Example 1 relates to the preparation and
structural analysis of cDNA for mouse Egr-1. Example 2
relates to confirmation of the presence of an Egr DNA
sequence on human chromosome 5. Example 3 relates to
the in vitro transcription and translation of mouse

- 9 -

Egr-1 cDNA. Example 4 relates to production of anti-
bodies according to the invention. Example 5 relates to
5 the isolation and characterization of genomic DNA which
encodes mouse Egr-1. Example 6 relates to the isolation
and characterization of cDNA encoding human EGR2.
10 Example 7 relates to preparation, in an E. coli host, of
a recombinant fusion protein including a portion of the
deduced amino acid sequence of mouse Egr-1. Example 8
relates to use of DNA probes of the invention in the
15 quantitative detection of EGR1 mRNA.

These examples are for illustrative purposes
only and are not intended in any way to limit the scope
20 of the invention.

<u>EXAMPLE 1</u>

25
Preparation and Structural
<u>Analysis of cDNA for Mouse Egr-1</u>

Isolation of DNA encoding a mammalian early
30 growth regulatory protein including one or more histi-
dine-cysteine zinc finger amino acid sequences was per-
formed substantially according to the procedures des-
cribed in Sukhatme et al., <u>Oncogene Research</u>, <u>1</u>, 343-355
35 (1987), the disclosures of which are specifically incor-
porated by reference herein.

Balb/c 3T3 cells (clone A31) from the American
Type Culture Collection were grown to confluence in
Dulbecco's Modified Eagle's medium (DME) supplemented
with 10% fetal calf serum (FCS). The cells were ren-
dered quiescent by reduction of the serum concentration
to 0.75% for 48 hours. To induce the cells from quies-
cence into growth phase G_1, the medium was changed to
20% FCS with cycloheximide added to a final concentra-
tion of 10 µg/ml.

RNA was extracted from Balb/c 3T3 cells har-
vested three hours after induction of quiescent cells by

20% FCS and 10 µg/ml cycloheximide. A λgt10 cDNA
library was constructed from this mRNA according to the

5   procedures of Huynh et al., DNA Cloning, Vol. 1, 49-78
(Glover, D., ed., IRL Press, 1985). This library was
screened differentially with single stranded cDNA pre-

10   pared from quiescent cells and from cells exposed to
serum and cycloheximide for 3 hours. These ^{32}P-labeled
cDNA probes were prepared from poly A$^+$ RNA as described
in St. John, et al., Cell, 16, 443-452 (1979), except

15   that 100 µCi of ^{32}P-dCTP (>3000 Ci/mmol), 0.02 mM cold
dCTP and 2-5 µg of poly A$^+$ RNA was used in each reac-
tion. The mean size of the reverse transcribed probes,

20   as assessed by alkaline agarose gel electrophoresis and
subsequent autoradiography, was about 700 bases.
Replica filter lifts (GeneScreenPlus, NEN-DuPont) were

25   prepared essentially as described by Benton et al.,
Science, 196, 180-192 (1977), and approximately 3 x 10^6
cpm of ^{32}P-cDNA were used per filter (90 mm diameter).
Hybridizations were carried out at 65°C in 1% SDS, 10%

30   dextran sulfate, and 1 M NaCl for a period of 16
hours. The filters were washed twice for twenty minutes
each time, first at room temperature in 2 x SSC

35   [Maniatis et al., Molecular Cloning, Cold Spring Harbor
Laboratory (New York, 1982)], then at 65°C in 2 x SSC,
1% NaDodSO$_4$ and finally at 65°C in 0.2 x SSC.
Autoradiograms were prepared by exposing the blots for
18 hours at -70°C with an intensifying screen.

A total of 10,000 cDNA clones from the Balb/c
3T3 λgt10 library were differentially screened.
Seventy-eight clones were found to hybridize preferen-
tially to single-stranded cDNA from fibroblasts stimu-
lated for 3 hours with 20% FCS and cycloheximide as
compared to single-stranded cDNA from quiescent cells.
Inserts from these clones were cross-hybridized to each
other, resulting in the sorting of forty clones into 7
cDNA families one of which was identified as c-fos.

Another cDNA clone, referred to as OC68, contained a
2.2 kb insert and was characterized further. This

5   insert was subcloned into the Eco RI site of pUC13 and
probes were generated for Northern blot analysis either
from the insert or the corresponding pUC plasmid.
Figure 1B illustrates a partial restriction digest map

10   of the OC68 clone ("R" representing restriction sites
for RsaI) along with that of a shorter clone, OC19t.
Two RsaI digestion fragments, derived from the 5' end of

15   clone OC68 and each comprising approximately 130 base
pairs, were labeled and employed to re-screen the above-
described λgt10 cDNA library, resulting in the recovery

20   of a 3.1 kb clone, designated OC3.1, shown in figure
1B. This clone was sequenced according to the method of
Sanger et al., Proc. Nat'l. Acad. Sci. (USA), 74, 5463

25   (1977). The 3086 base pair sequence obtained is set
forth in Figure 1A along with the deduced sequence of
533 amino acid residues for the protein encoded,
designated mouse "Egr-1".

30   The deduced amino acid sequence shows a single
long open reading frame with a stop codon (TAA) at posi-
tion 1858. The most 5', in-frame, ATG, at position 259,

35   is flanked by sequences that fulfill the Kozak criterion
(^A_GNN(ATG)G) [Kozak, Nuc. Acids Res., 15, 8125-8131
(1987)]. The sequence region upstream of this ATG is
highly GC-rich and results in an absence of in-frame
stop codons. The 3' untranslated region (UT) contains
two "AT" rich regions (nucleotides 2550-2630 and
2930-2970). Similar sequences are found in the 3' UT
regions of several lymphokine and proto-oncogene mRNAs,
including granulocyte macrophage colony stimulating
factor (GM-CSF), interleukin 1, interleukin 2, inter-
leukin 3 (IL-3), α, β, and γ interferons, and c-fos,
c-myc, and c-myb [Shaw et al., Cell, 46, 659-667
(1986)]. These sequences may mediate selective mRNA
degradation. The presence in the mouse Egr-1 transcript

of such regions is consistent with its short message

half-life. Potential polyadenylation signals (AATAAA)

5   are located at nucleotide positions 1865 and 3066, as

well as at position 3053 (AATTAA) [Wickens et al.,

Science, 226, 1045-1051 (1984)].

10   The deduced amino acid sequence predicts a

polypeptide of 533 amino acids with a molecular weight

of 56,596. Based on structural considerations, namely a

central region containing zinc fingers (described

15 below), the Egr-1 protein can be divided into three

domains. The N-terminal portion (amino acid residues 2

to 331) is rich in proline (14.2%) and serine (16%)

20 residues with 7.9% alanines and 7.9% threonines. The C-

terminal region (residues 417 to 533) also contains a

very high proportion of prolines and serines (15.4 and

25 26.5%, respectively) and 10.3% alanines and 11.1%

threonines. The large number of proline residues leads

to a secondary structure that probably lacks

α-helices. The central portion of the Egr-1 protein

30 consists of three tandem repeat units of 28-30 amino

acids, with the first unit starting at position 332.

Each unit conforms almost exactly to the consensus

35 sequence $TGX_{3F}^{Y}XCX_{2-4}CX_3FX_5LX_2HX_3H$ (see Figure 2),

diagnostic of DNA binding zinc fingers [Berg, Science,

232, 485-486 (1986); Brown et al., Nature, 324, 215

(1986); and Brown et al., FEBS Letters, 186, 271-274

(1985)]. Furthermore, the Egr-1 fingers are connected

by "H-C links" ($TGE_{K}^{R}P_{Y}^{F}X$) [Schuh et al., Cell, 47,

1025-1032 (1986)] found in the Xenopus TFIIIA gene

(between fingers 1, 2, and 3), in the Drosophila Kruppel

gap gene [Rosenberg et al., Nature, 319, 336-339

(1986)], and in genes from mouse and Xenopus that cross-

hybridize to the Kruppel (Kr) finger domains: mkr1,

mkr2 [Chowdhury et al., Cell, 48, 771-778 (1987)], and

Xfin [Altaba et al., EMBO Journal, 6, 3065-3070

(1987)]. The sequence similarity amongst the Egr-1

- 13 -

fingers is 50-70%, whereas the sequence similarity
between any of the Egr-1 fingers and those present in
TFIIIA, Kruppel, mkr1, mkr2 or Xfin is 35-40%. Outside
of the finger domains, it is noteworthy that the Egr-1
and Kr proteins each contain a very high proportion of
Pro, Ala, and Ser residues [Schuh et al., Cell, 47,
1025-1032 (1986)]. However, there is no sequence
similarity in these regions. Thus, Egr-1 and Kr are not
homologous genes nor is Egr-1 related to mkr1, mkr2,
Xfin, or TFIIIA. The Kr gene contains thirteen copies
of the hexanucleotide (ACAAAA), or its complementary
sequence, eight of which are located within 180 bp
downstream from the Kr TATA box and five are in the 3'
UT region. These sequences may serve as targets for
other DNA binding proteins or in Kr gene autoregula-
tion. The Egr-1 cDNA also contains nine copies of the
ACAAAA sequence or its complement.

Following the work described above, Milbrandt
[Science, 238, 797-799 (1987)], reported the isolation
and sequence of a nerve growth factor (NGF) inducible
cDNA (NGFI-A) from the rat pheochromocytoma PC12 line.
A comparison of the deduced amino acid sequence of
NGFI-A to that of mouse Egr-1 of Figure 1A reveals 98%
sequence identity. Thus, mouse Egr-1 and rat NGFI-A are
homologs. The putative initiation ATG chosen by
Milbrandt corresponds to position 343 in the Figure 1A
cDNA sequence, and is 84 nucleotides (28 amino acid
residues) downstream of the ATG therein designated for
translation initiation. Both ATG's have a purine at
position -3 and a G at position +1 and the designation
represented in Figure 1A of the more 5' ATG as the puta-
tive start codon is based on the experience of Kozak,
Nuc. Acids Res., 15, 8125-8131 (1987), even though the
more 3' ATG is surrounded by the longer Kozak consensus
sequence (CCG/ACCATGG). Translation of an in vitro
generated RNA transcript, described infra, selects the
more 5' ATG for initiation.

- 14 -

It is noteworthy that a major difference in the deduced sequences of mouse Egr-1 and rat NGFI-A resides in the sequence spanning residues 61-68 of Egr-1 and 33-43 of NGFI-A. The former includes the sequence N-S-S-S-S-T-S-S while the latter includes the sequence N-N-S-S-S-S-S-S-S-S-S, accounting for the 3 residue difference in length of the putative polypeptides which is not accounted for by the difference in designation of the transcript initiation signal.

EXAMPLE 2

Human Chromosome Gene Mapping

To determine the human chromosomal localization of the gene corresponding to mouse Egr-1, the OC3.1 and OC19t cDNA clones were hybridized to a panel of rodent x human somatic cell hybrids. Southern blot analysis of the hybrid panel showed concordance between the presence of Egr-1 sequences and human chromosome 5. In situ hybridization to normal human metaphase chromosomes resulted in specific labeling only of chromosome 5, with the largest cluster of grains at 5q23-31. Specific labeling of these bands was also observed in hybridizations using an Egr-1 probe which does not contain finger sequences.

This localization is interesting in light of the non-random deletions [del(5q)] in human myeloid disorders (acute myelogenous leukemia) (AML), and myelodysplastic syndromes, that involve this chromosomal region. [Le Beau et al., Science, 231, 984-987 (1986); Dewald et al., Blood, 66, 189-197 (1985); and Van den Berghe et al., Cancer Genet. Cytogenet., 17, 189-255 (1985)]. Fifty percent of patients with therapy related AML show chromosome 5 abnormalities (interstitial deletions or monosomy) and cytogenetic analysis of the deletions has revealed that one segment, consisting of bands

- 15 -

5q23-31, is absent in the malignant cells of all
patients who have aberrations of chromosome 5. These
5   data suggest that loss of a critical DNA sequence lead-
ing to hemizygosity (or homozygosity) of a recessive
allele may play an important role in the pathogenesis of
these disorders, a mechanism substantiated for retino-
10 blastoma. Although genes for a number of growth factors
and receptors (IL-3, GM-CSF, β_2-adrenergic receptor,
endothelial cell growth factor, CSF-1, c-fms, pDGF
15 receptor) are clustered in or near this region, Egr-1
(by virtue of its zinc fingers) is the only member of
this group with potential transcriptional regulatory
20 activity. It is therefore possible that its absence
could lead to deregulated cell growth.

EXAMPLE 3

25

In Vitro Expression of Mouse Egr-1 cDNA
            A 2.1 kb ApaI/ApaI fragment (comprising
30 nucleotides 120-2224 of Figure 1A) was isolated from the
OC3.1 DNA clone. This fragment includes the translation
start (ATG) codon at nucleotide position 259 designated
in Figure 1A. The fragment was blunt-ended with T4 DNA
35 polymerase and cloned into the Bluescript vector KS
M13(+) containing a T3/T7 bacteriophage promoter. The
(T3) sense transcript was generated and in vitro trans-
lated in a standard rabbit reticulocyte lysate system
(Promega Biotec, Madison, WI. 53711) including
^{35}S methionine as a radiolabel. An analogous in vitro
transcription system was developed using a BglII/BglII
fragment of OC3.1 (including nucleotides 301-1958 and
not including the translation start designated in Figure
1A). The T7 sense transcript was employed in the trans-
lation system. Differential characterization of trans-
lation products by autoradiographic SDS PAGE indicated
that the ATG at nucleotide position 259 is preferred as

- 16 -

a translation start codon when all potential start sites
are present.

EXAMPLE 4

Preparation of Antibodies:
        A first synthetic peptide based on the
sequence of amino acid residues 416-427 of mouse Egr-1
was prepared and provided with a carboxy terminal
cysteine residue. The peptide, H-L-R-Q-K-D-K-K-A-D-K-
S-C, was coupled to KLH and employed to immunize New
Zealand white rabbits. Animals were initially immunized
with 100 µg of the immunogen in Freund's Complete
Adjuvant and every two weeks were boosted with 100 µg of
immunogen with Freund's Incomplete Adjuvant. Sera,
designated VPS10, were isolated after 68 days and
displayed an antibody titer of 1:12,800 based on
reactivity with the antigen used to prepare the anti-
sera.
        A second synthetic peptide, based on residues
399 to 415 of mouse Egr-1, was prepared. The peptide,
C-G-R-K-R-A-R-S-D-E-R-K-R-H-T-K-I, was coupled to KLH
and used to immunize rabbits as above, resulting in the
production of antisera (designated VPS2) with a titer of
1:400.

EXAMPLE 5

Isolation of Genomic Mouse Egr-1 Clone and
Characterization of Regulatory Regions
        A mouse Balb/c 3T3 genomic library was pre-
pared in a Stratagene (La Jolla, California) vector,
λFIX, according to the manufacturer's instructions and
probed using 1% SDS, 1 M NaCl, and 10% dextran sulfate
at 65°C with stringent final wash in 0.2 x SSC at 65°C
with a 2.1 kb ApaI/ApaI fragment and a 3.1 kb Eco RI/Eco

RI fragment derived from digestion of pUC13 including the mouse Egr-1 clone OC3.1. One positive clone, from approximately 300,000 screened, was designated mgEgr-1.1 and also hybridized to the extreme 5'-end 120 bp Eco RI-Apa I fragment from plasmid OC3.1.

A 2.4 kb Pvu-II-PvuII fragment and a 6.6 kb XbaI-XbaI fragment (see Figure 5) derived from the mgEgr-1.1 clone were subcloned into the SmaI and XbaI sites of pUC13 and pUC18 respectively, and the resulting plasmids (designated as p2.4 and p6.6) were used for restriction mapping analysis of transcription initiation sites and for nucleotide sequencing. Marked in Figure 4, and listed in Table 1, are possible regulatory elements identified in the 5' flanking sequence of mgEgr-1.1. A putative TATA motif (AAATA) is located 26 nucleotides upstream of the transcription start site. A "CCAAT" type sequence starts at nucleotide -337. Five different regions, each 10 nucleotides in length, located at -110, -342, -358, -374, and -412, are nearly identical to the inner core of the c-fos serum response element (Treisman, R., Cell, 46, 567 (1986)). Each has a 5-6 nucleotide AT rich stretch and is surrounded by the dinucleotide CC on the 5' side and GG on the other. Two potential TPA responsive elements (Lee, W., et al., Cell, 49, 741 (1987) and Angel, P., et al., Cell, 49, 729 (1987)) are located at nucleotides -610 and -867. Four consensus Spl (Briggs, M.R., et al., Science, 234, 47 (1986) binding sequences are at position -285, -649, -700 and -719. In addition, two sequences have been identified that might serve as cAMP response elements (Montimy, M.R., et al., Nature, 328, 175 (1987)) (-138 and -631).

- 18 -

TABLE 1

Location and Identification of
Potential Regulatory Elements

Element	Sequence[1]	Location[2]
TATA	AAATA	-26 to -22
CCAAT	CCAAT	-337 to -333

Serum Response Element
Consensus

GATGTCCATATTAGGACATC	TCCTTCCATATTAGGGCTTC	-110 to -91
CC TA AT GG	GTGGCCC-AATATGGCCCTG	-342 to -324
G C	CAGCGCCTTATATGGAGTGG	-358 to -339
	ACAGACCTTATTTGGGCAGC	-374 to -355
	AAACGCCATATAAGGAGCAG	-412 to -393

TPA Responsive Element
(AP1 binding site)
Consensus

C C	CTGACTCG	-610 to -603
TGACT A	CTGACTGG	-867 to -860
G A		

Sp1 binding site	GGGCGG	-285 to -280
	GGGCGG	-649 to -644
	CCGCCC	-700 to -695
	GGGCGG	-719 to -714

cAMP Response Element
Consensus

TGACGTCA	TCACGTCA	-138 to -131
	TGACGGCT	-631 to -624

---

1. The underlined bases in the mouse Egr-1 gene
sequence are those that do not match the consensus
sequence.

2. The location numbers refer to the nucleotides of the
mouse Egr-1 gene as indicated in Figure 4.

- 19 -

To obtain the genomic sequence and the intron-exon gene structure, specific oligonucleotides (17-mers at positions 83, 122, 174, 200, 379, 543, 611, 659, 905, 920, 1000, 1200, 1400, 1600, 1800, 2100, 2353, 2650, 2825) of the OC3.1 cDNA sequence (see Figure 1A) were used as primers for double stranded sequencing of plasmids p2.4 and p6.6. Comparison of the Egr-1 genomic sequence to the Egr-1 cDNA sequence showed the Egr-1 gene consists of 2 exons and a single 700 bp intron (between nucleotide position 556 and 557 as numbered in Figure 1A and as shown in Figure 5). Both the 5' and 3' splice junction sequences (not shown) are in excellent agreement with the consensus boundary sequences. Mount, S.M., Nucleic Acids Res., 10, 459 (1982).

EXAMPLE 6

Isolation and Characterization
of Human EGR2 cDNA

A human genomic placental library in the vector EMBL3, prepared by Dr. C. Westbrook of the University of Chicago according to procedures described in Frischauff et al., Jour. Mol. Biol., 170, 827-842 (1983), and a human leukocyte cosmid library prepared according to procedures described in Proc. Nat'l. Acad. Sci. (USA), 80, 5225-5229 (1983), were probed with the 2.1 kb ApaI fragment of OC3.1 (described in Example 5) using 1% SDS, 1 M NaCl and 10% dextrose sulfate at 50-55°C with a non-stringent final wash in 2 x SSC at 50-55°C. A single positive clone (designated HG6) was isolated from the first library and four clones (designated HG17, 18, 19 and 21, respectively) were isolated from the second library. A 6.6 kb SalI/EcoRI fragment of clone HG6 was found to hybridize with a 332 base pair HpaII/HpaII fragment of the mouse Egr-1 gene,

- 20 -

which letter fragment spans the putative zinc finger
region. The 6.6 kb fragment, in turn, was employed to
5   probe a cDNA library derived from human fibroblasts
which have been stimulated for three hours with 20%
fetal calf serum in the presence of 10 µg/ml
cyclohexamide. About 10,000 clones were screened and
10   the fifty positive clones obtained (designated "zap-1
through zap-50") are being subjected to nucleotide
sequence analysis. Preliminary sequence analysis
15   reveals that three clones, zap-2, zap-8, and zap-32, all
encode the same transcript, namely a protein designated
human EGR2, shown in Figure 3. Preliminary analysis
20   indicates approximately 92% homology between mouse Egr-1
and human EGR2 polypeptides in the zinc finger regions,
but substantially less homology in the amino and carboxy
terminal regions. Chromosome mapping studies, similar
25   to those described in Example 2, indicate that human
chromosome 10, at bands q21-22, constitutes a locus for
the human EGR2 gene.
30   The plasmid zap-32, containing the full length
human EGR2 clone, was used as a probe in Southern blot
analysis on DNAs from 58 unrelated Caucasians. It was
35   found that Hind III detects a simple two-allele
polymorphism with bands at either 8.0 kb (A1) or 5.6 kb
and 2.4 kb (A2). No constant bands were detected. The
frequency of A1 was 0.90 and that of A2 was 0.10. No
polymorphisms were detected for Apa I, BamH I, Ban II,
Bgl I, Bgl II, BstE II, Dra I, EcoR I, EcoR V, Hinc II,
Msp I, Pst I, Pvu II, Rsa I, Sac I, and Taq I in 10
unrelated individuals. Co-dominant segregation of the
Hind III RFLP was observed in four large kindreds with a
total of more than 350 individuals.

           These data will be useful in gene linkage
studies for mapping genes for certain genetic
disorders. For example, the gene responsible for the
dominantly inherited syndrome, multiple endocrine

neoplasia, type 2A (MEN-2A) has been assigned by linkage
to chromosome 10. Simpson, et al., Nature, 328, 528
5   (1987). Studies are currently underway to determine the
linkage relationship between MEN-2A and EGR2 and are
expected to be useful in cloning the MEN-2A gene as well
as in serving as a diagnostic marker for the disease.
10

EXAMPLE 7

15   Recombinant Expression Of Fusion Proteins
          A 322 base HpaII/HpaII fragment (comprising
nucleotides 1231-1553) derived from the OC3.1 cloned DNA
20   was treated with DNA polymerase to fill in the single
stranded ends. This fragment was inserted in plasmid
pEX3 (obtained from K. Stanley, European Molecular
Biology Laboratory, Postfach 10.2209, 6900 Heidelberg,
25   F.R.G.) digested with SmaI. Stanley, K.K., et al., EMBO
J., 3, 1429 (1984). This insertion placed the Egr-1
encoding DNA fragment in the same reading frame as
30   plasmid DNA encoding cro-β-galactosidase, allowing for
the expression of a fusion protein comprising the amino
terminal residues of cro-β-galactosidase and 108 resi-
35   dues of Egr-1 amino acids 325 to 432. This cro-β-
galactosidase/Egr-1 fusion plasmid, designated pFIG, was
used to transform E. coli NF1.
          Induced (42°C) and un-induced (30°C) cultured
cell lysates from growth of the transformed NF1 cells
were then analyzed by SDS-PAGE. Upon Coomassie stain
analysis, only induced cell lysates included an
approximately 108 kd product, indicating presence of the
projected expression product. Western blot analysis,
using the rabbit polyclonal anti-peptide antibody VPS10
(see Example 4) raised against H-L-R-Q-K-D-K-K-A-D-K-S-
C, confirmed that the fusion protein product contained
Egr sequences.

- 22 -

In a separate construction, a mouse Egr-1
insert, from plasmid OC3.1, was fused, in frame, to a
plasmid containing sequences from bovine growth hormone
according to the methods described in Slamon, D.J., et
al., Science, 233, 347 (1986). The resultant plasmid,
designated pV4, comprised a fusion protein containing a
fusion gene coding for bovine growth hormone amino acids
1 to 192 and Egr-1 amino acids 2 to 533. This bGH/mouse
Egr-1 DNA fusion plasmid, designated pV4, was expressed
in E. coli and the resulting fusion protein, designated
V4, was identified in Western blots by its reactivity
with a bGH monoclonal antibody and its reactivity with
VPS10 rabbit anti-Egr-1 peptide antiserum, prepared
according to Example 4.

EXAMPLE 8

Determination of Egr
Levels in Human Tumor and
Non-Tumor Tissue

Using the mouse Egr-1 OC68 probe, Northern
blot analyses were conducted to determine the levels of
transcription of Egr protein encoding DNA in tumor
versus surrounding normal tissue from resected human
tumor specimens. The tumor samples were from lung (12),
colon (7), colon mesastasis (1), bladder (1), rectal
(1), giant cell (1), hepatoma (1), breast (1), MFH
(malignant fibrous histiocytoma) (1), osteosarcoma (1)
and rhabdomyosarcoma (1). In about 50% of these cases,
there is markedly decreased (about three to ten-fold)
expression of the Egr mRNA in tumor versus normal
tissue. One implication of this finding is that Egr
proteins of the invention may function as part of a
negative regulatory pathway. In any event, it is clear
that DNA sequences and antibodies of the invention are
susceptible to use in differential diagnoses between
tumorous and non-tumorous cell types.

It will be apparent from consideration of the
foregoing illustrative examples that the present inven-
tion constitutes a substantial advance in the art and
the achievement of a major goal in molecular biology,
i.e., the characterization of genes which play a regu-
latory role in mammalian cell proliferation and differ-
entiation. It will thus be understood that the informa-
tion provided herein constitutes a basis for straight-
forward development of useful methods and materials not
specifically the subject of the above examples. By way
of illustration, possession of knowledge concerning the
base sequence of cDNA and genomic DNA sequences encoding
distinct mouse Egr-1 and human EGR2 early growth regu-
latory proteins comprising histidine-cysteine zinc
finger amino acid sequences makes possible the isolation
of other such structurally related proteins. The sub-
stantial homology between the zinc finger regions of
Egr-1 and EGR2 coupled with lack of homology in other
protein regions, when considered in light of the ability
of Egr-1 probes to localize to human chromosome 5 while
EGR2 probes localize to human chromosome 10, essentially
assures the straightforward isolation of a human gene
(provisionally designated "human EGR1") which encodes a
protein more closely homologous to Egr-1 and a mouse
gene (Egr-2) encoding a protein more closely homologous
to EGR2.

While the above examples provide only limited
illustration of _in vitro_ and _in vivo_ expression of DNA
sequences of the invention, known recombinant techniques
are readily applicable to development of a variety of
procaryotic and eucaryotic expression systems for the
large scale production of Egr proteins and even develop-
ment of gene therapy regimens.

Knowledge of the specifically illustrated
mouse Egr-1 and human EGR2 proteins of the invention has
been demonstrated to provide a basis for preparation of

- 24 -

highly useful antibodies, also provides a wealth of
information concerning the nature of protein-nucleic

5   acid interactions which, in turn, constitutes a basis
for determination of significant early growth regulatory
events. For example, and by analogy to steroid receptor

10 protein structures, analysis of the structure of regions
flanking the zinc fingers of Egr-1 and EGR2 and related
proteins of the invention is expected to allow for
identification of substances which may interact with the

15 proteins to alter their DNA interactive capacities and
thus provide the basis for inhibition or augmentation of
their regulatory functions. Moreover, information

20 available concerning specific events of DNA interaction
of Egr proteins of the invention will permit, e.g.,
identification and use of potential competitive

25 inhibitors of these proteins.
        Just as Egr encoding DNA of the invention is
conspicuously susceptible to use in differentiation of
human tumor and non-tumor cells, antibodies prepared

30 according to the invention are expected to be useful in
differential screening of cells based on relative
cellular concentrations of mRNA expression products and

35 in the determination of specific genes susceptible to
regulation by such products.
        Because numerous modifications and variations
in the practice of the present invention are expected to
occur to those skilled in the art, only such limitations
as appear in the appended claims should be placed
thereon.

- 25 -

WHAT IS CLAIMED IS

1. A purified and isolated DNA sequence encoding a mammalian early growth regulatory protein which comprises one or more histidine-cysteine zinc finger amino acid sequences.

2. The DNA sequence according to claim 1 encoding human EGR2 protein.

3. The DNA sequence according to claim 1 encoding mouse Egr-1 protein.

4. The DNA sequence according to claim 1 which is a cDNA sequence.

5. The DNA sequence according to claim 1 which is a genomic DNA sequence.

6. The DNA sequence according to claim 1 which is a wholly or partially synthetic DNA sequence.

7. The DNA sequence according to claim 1 operatively associated with an homologous or heterologous expression control DNA sequence.

8. The DNA sequence according to claim 1 selected from the group consisting of the DNA sequences set out in Figures 1A, 3, and 4.

9. A procaryotic or eucaryotic host cell transformed or transfected with a DNA sequence comprising a DNA sequence according to claim 1 or 7.

10. A viral or circular DNA plasmid vector comprising a DNA sequence according to claim 1.

- 26 -

5   11.   The viral or circular DNA plasmid vector
according to claim 10 further comprising an expression
control DNA sequence operatively associated with said
early growth regulatory protein encoding DNA.

10

12.   A method for the production of an early
growth regulatory protein comprising:
           growing, in culture, a host cell transformed
15 or transfected with a DNA sequence according to claim 1;
and
           isolating from said host cell or culture the
20 polypeptide product of the expression of said DNA
sequence.

25   13.   A method for the production of an early
growth regulatory protein comprising:
           disposing a DNA sequence according to claim 1
in a cell free transcription and translation system; and
30   isolating from said system the polypeptide
product of the expression of said DNA sequence.

35   14.   The polypeptide product of the in vitro
or in vivo expression of part or all of a protein encoding
region of a DNA sequence according to claim 1.

15.   The polypeptide product according to
claim 14 which is a fusion protein comprising part or
all of a mammalian early growth regulatory protein which
comprises one or more histidine-cysteine zinc finger
amino acid sequences and part or all of a heterologous
proteins.

16.   The polypeptide product according to
claim 15, which comprises a fusion of cro-β-
galactosidase and Egr-1 amino acid sequences or bovine
growth hormone and Egr-1 amino acid sequences.

- 27 -

17.   A purified and isolated mammalian early
growth regulatory protein comprising one or more histi-
dine-cysteine zinc finger amino acid sequences.

18.   The protein according to claim 17 and
having the amino acid sequence set out in Figure 1A.

19.   The protein according to claim 17 and
having the amino acid sequence set out in Figure 3.

20.   A synthetic peptide duplicative of a
sequence of amino acids present in a mammalian early
growth regulatory protein comprising one or more histi-
dine-cysteine zinc finger amino acid sequences and
sharing at least one antigenic epitope of such
protein.

21.   The synthetic peptide of claim 20 and
having an amino acid sequence partially duplicative of
the amino acid sequence set out in Figure 1A.

22.   The synthetic peptide of claim 20 and
having an amino acid sequence partially duplicative of
the amino acid sequence set out in Figure 3.

23.   The synthetic peptide of claim 20 and
further characterized as duplicative of an amino acid
sequence not involved in DNA binding functions.

24.. An antibody specifically immunoreactive
with at least one epitope of a mammalian early growth
regulatory protein comprising one or more histidine-
cysteine zinc finger amino acid sequences.

- 28 -

25.   The antibody according to claim 24
wherein said epitope is other than an epitope within the
DNA binding functional region thereof.

26.   The antibody according to claim 24
selected from the group consisting of monoclonal and
polyclonal antibodies.

27.   The antibody according to claim 24
capable of specifically binding with a proteinaceous
material comprising an amino acid sequence duplicating
an antigenic epitope within the following amino acid
sequence:
H-L-R-Q-K-D-K-K-A-D-K-S-C.

28.   The antibody according to claim 24
capable of specifically binding with a proteinaceous
material comprising an amino acid sequence duplicating
an antigenic epitope within the following amino acid
sequence:
C-G-R-K-F-A-R-S-D-E-R-K-R-H-T-K-I.

29.   A method for quantitative detection of a
mammalian early growth regulatory protein comprising one
or more histidine-cysteine zinc finger amino acid
sequences based on the immunological reaction of the
same with an antibody according to claim 24.

30.   A method for quantitative detection
within a sample of messenger RNA transcripts for
mammalian early growth regulatory proteins comprising
one or more histidine-cysteine zinc finger amino acid
sequences comprising the step of hybridizing RNA within
said sample with a DNA sequence partially or wholly
duplicating a DNA sequence according to claim 1.

31. A method for quantitative detection within a sample of DNA encoding for mammalian early growth regulatory proteins comprising one or more histidine-cysteine zinc finger amino acid sequences comprising the step of hybridizing DNA within said sample with a DNA sequence partially or wholly duplicating a DNA sequence according to claim 1.

32. A method for detecting a disease genetically linked to a mammalian Egr gene comprising the step of quantitating mammalian early growth regulatory DNA sequences according to claim 1.

```
        10        20        30        40        50        60
GGGGAGCCGCCGCCGCGATTCGCCGCCGCCCGCCGCTTCCGCCGCCCGCCAAGATCGGCCCC

        70        80        90       100       110       120
TGCCCCAGCCTCCGGGCCAGCCCGGCGTCCACCACGGCCCGCGGCTACCGGCCAGCCTGGG

       130       140       150       160       170       180
GGCCCACCTACACTCCCCGCAGTGTGCCCCTGCACCCCGCATGTAACCGGCCAACCCCC

       190       200       210       220       230       240
GGCGAGTGTGCCCTCAGTAGCTTCGGGCCCGGGCTGCGCCCACCACCCAACATCAGTTCT

       250       260       270       280       290       300
CCAGCTCGCTGGTCCGGGATGGCAGCGGCCAAGGCCGAGATGCAATTGATGTCTCCGCTG
                         MetAlaAlaAlaLysAlaGluMetGlnLeuMetSerProLeu

       310       320       330       340       350       360
CAGATCTCTGACCCGTTCGGCTCCTTCCTCACTCACCACCATGGACAACTACCCCAAA
GlnIleSerAspProPheGlySerPheProThrMetAspAsnTyrProLys

       370       380       390       400       410       420
CTGGAGGAGATGATGCTGCTGAGCAACGGGGCTCCCAGTTCCTCGGTGCTGCCGGAACC
LeuGluGluMetMetLeuLeuSerAsnGlyAlaProGlnPheLeuGlyAlaAlaAlaGlyThr
```

FIGURE 1.1

```
           430       440       450       460       470       480
CCAGAGGGCAGCGGCGGTAATAGCAGCAGCAGCACCAGCAGCGGCGGGCGGTGGTGGGGGC
ProGluGlySerGlyGlyAsnSerSerSerThrSerSerGlyGlyGlyGlyGlyGly

           490       500       510       520       530       540
GGCAGCAACAGCGGGCAGCAGCGCCTTCAATCCTCAAGGGGAGCCGAGCGAACAACCCTAT
GlySerAsnSerGlySerSerAlaPheAsnProGlnGlyGluProSerGluGlnProTyr

           550       560       570       580       590       600
GAGCACCTGACCACCAGAGTCCTTTCTGACATGCTCTGAATAATGAGAAGGCGATGGTG
GluHisLeuThrThrGluSerPheSerAspIleAlaLeuAsnAsnGluLysAlaMetVal

           610       620       630       640       650       660
GAGACGAGTTATCCCAGCCAAACGACTCGGTTGCCTCCCATCACCTATACTGGCCGCTTC
GluThrSerTyrProSerGlnThrThrArgLeuProProIleThrTyrThrGlyArgPhe

           670       680       690       700       710       720
TCCCTGGAGCCCGCACCAACAGTGGCAACACTTGTGGCCTGAACCCCTTTCAGCCTA
SerLeuGluProAlaProAsnSerGlyAsnThrLeuTrpProGluProLeuPheSerLeu

           730       740       750       760       770       780
GTCAGTGGCCTCGTGAGCATGACCAATCCTCCGACCTCTTCAATCCTCCGGCCCTTCTCCA
ValSerGlyLeuValSerMetThrAsnProProThrSerSerSerAlaProSerPro

           790       800       810       820       830       840
GCTGCTTCATGCGTTCTTCCTCTGCCTCCCAGAGCCCCGAGCTGTGCCGTGCCGTCC
AlaAlaSerSerSerAlaSerGlnSerProLeuSerProLeuSerCysAlaValProSer
```

FIGURE 1.2

SUBSTITUTE SHEET

```
        850       860       870       880       890       900
AACGACAGCAGTCCCATCTACTCGGCTGCGCCCACCTTCCTACTCCCAACACTGACATT
AsnAspSerProIleTyrSerAlaAlaProThrPheProThrProAsnThrAspIle

        910       920       930       940       950       960
TTTCCTGAGCCCCAAAGCCAGGCCTTTCCTGGCTCGGCAGGCACAGCCTTGCAGTACCCG
PheProGluProGlnSerGlnAlaPheProGlySerAlaGlyThrAlaLeuGlnTyrPro

        970       980       990      1000      1010      1020
CCTCCTGCTACCCTGCCACCAAAGGTGGTTCCCAGGTTCCCATGATCCCTGACTATCTG
ProProAlaThrLeuProProLysValValProArgPheProMetIleProAspTyrLeu

       1030      1040      1050      1060      1070      1080
TTTCCACAACAGGAGGAGACCTGAGCCTGGGCACCCAGACCAGAAGCCCTTCCAGGGT
PheProGlnGlnGluGluThrLeuSerLeuGlyThrProAspGlnLysProPheGlnGly

       1090      1100      1110      1120      1130      1140
CTGGAGAACCGTACCCAGCAGCCTTCGCTCACTCCACTATTCCACTATTAAAGCCTTCGCC
LeuGluAsnArgThrGlnGlnProSerLeuThrProLeuSerThrIleLysAlaPheAla

       1150      1160      1170      1180      1190      1200
ACTCAGTCGGGCTCCCAGGACTTAAAGGCTCTTAATACCACCTACCACCAATCCCAGCTCATC
ThrGlnSerGlySerGlnAspLeuLysAlaLeuAsnThrThrTyrGlnSerGlnLeuIle

       1210      1220      1230      1240      1250      1260
AAACCCAGCCGCATGCGCAAGTACCCAACCGGCCCAGCAAGACACCCCCCATGAACGC
LysProSerArgMetArgLysTyrProAsnArgProSerLysThrProProHisGluArg
```

SUBSTITUTE SHEET

FIGURE 1.3

```
      1270      1280      1290      1300      1310      1320
CCATATGCTTGCCCTGTCGAGTCCTGCGATGCCGCTTTCTGCTCGGATGAGCTTACC
ProTyrAlaCysProValGluSerCysAspArgArgPheSerArgSerAspGluLeuThr

      1330      1340      1350      1360      1370      1380
CGCCATATCCGCATCCACACAGGCCAGGAAGCCCTTCCAGTGTGCAATCTGCATGCGTAAC
ArgHisIleArgIleHisThrGlyGlnLysProPheGlnCysArgIleCysMetArgAsn

      1390      1400      1410      1420      1430      1440
TTCAGTCGTAGTGACCACCTTACCACCCACAATCCGCACCACCAGGAGAAGCCTTTT
PheSerArgSerAspHisLeuThrThrHisIleArgThrHisThrGlyGluLysProPhe

      1450      1460      1470      1480      1490      1500
GCCTGTGACATTTGTGGGGAGGAAGTTTGCCAGGAGTGATGAACGCAAGAGAGGCATACCAAA
AlaCysAspIleCysGlyArgLysPheAlaArgSerAspGluArgLysArgHisThrLys

      1510      1520      1530      1540      1550      1560
ATCCATTTAAGACAGAAGGACAAGAAAGCAGACAAAAGTGTGGTGGCCGTCCCCGGCTGCC
IleHisLeuArgGlnLysAspLysLysAlaAspLysSerValValAlaSerProAlaAla

      1570      1580      1590      1600      1610      1620
TCTTCACTCTCTTCTTACCCATCCCCAGTGGCTCCTACCCATCCCCTGCCACCACC
SerSerLeuSerSerTyrProSerProValAlaThrSerTyrProSerProAlaThrThr

      1630      1640      1650      1660      1670      1680
TCATTCCCATCCCCTGTGCCCACTTCCTACTCCTCCTGGCTCCTCCACCTACCCATCT
SerPheProSerProValProThrSerSerSerProGlySerSerProSerThrTyrProSer
```

FIGURE 1.4

SUBSTITUTE SHEET

```
                1690      1700      1710      1720      1730      1740
CCTGGCGCAGTGGCTTCCCGTCGGCCGTGCCCAGTGGCCACCACCTTTGCCTCCGTTCCACCT
ProAlaHisSerGlyPheProSerProSerValAlaThrThrPheAlaSerValProPro

                1750      1760      1770      1780      1790      1800
GCTTTCCCCACCCAGGTCAGCAGCTTCCCGTCTGCGGGCGTCAGCAGCTCCTTCAGCACC
AlaPheProThrGlnValSerSerPheProSerAlaGlyValSerSerSerPheSerThr

                1810      1820      1830      1840      1850      1860
TCAACTGGTCTTTCAGACACATGACAGCGACCTTTCTCCCAGGACAATTGAAATTGCTAA
SerThrGlyLeuSerAspMetThrAlaThrPheSerProArgThrIleGluIleCys

                1870      1880      1890      1900      1910      1920
AGGGAATAAAAGAAAGCAAAGGGAGAGGCAGGAAAGACATAAAAGCACAGGAGGGAAGAG

                1930      1940      1950      1960      1970      1980
ATGGCCGCAAGAGGGCCACCTCTTAGGTCAGATGGAAGATCTCAGAGCCAAGTCCTTCT

                1990      2000      2010      2020      2030      2040
ACTCACGAGTAGAAGGACCGTTGGCCAACAGCCCTTCACTTACCATCCCTGCCCCCCC

                2050      2060      2070      2080      2090      2100
GTCCTGTTCCCTTTGACTTCAGCTGCCTGAAACAGCCATGTCCAAGTTCTTCACCTCTAT
```

FIGURE 1.5

```
      2110      2120      2130      2140      2150      2160
CCAAAGGACTTGATTTGCATGGTATTGGATAAATCATTTCAGTATCCTCTCCATCACATG

      2170      2180      2190      2200      2210      2220
CCTGGCCCTTGCTCCCTTCAGCGGCTAGACCATCAAGTTGGCATAAAGAAAAAAATGGG

      2230      2240      2250      2260      2270      2280
TTTGGGCCCTCAGAACCCTGCCCTGCCATCTTTGTACAGCATCTGTGCCATGGATTTGTT

      2290      2300      2310      2320      2330      2340
TTCCTTGGGGTATTCTTGATGTGAAGATAATTTGCATACTCTATTGTATTATTTGGAGTT

      2350      2360      2370      2380      2390      2400
AAATCCTCACTTTGGGGGAGGGGGAGCAAAGCCAAGCAAACCAATGATGATCCCTCTATT

      2410      2420      2430      2440      2450      2460
TTGTGATGACTCTGCTGTGACATTAGGTTTGAAGCATTTTTTTTTTCAAGCAGCAGTCCT

      2470      2480      2490      2500      2510      2520
AGGTATTAACTGGAGCATGTGTCAGAGTGTTGTTCCGTTAATTTGTAAATACTGGCTCG
```

FIGURE 1.6

SUBSTITUTE SHEET

```
          2530      2540      2550      2560      2570      2580
ACTGTAACTCTCACATGTGACAAAGTATGGTTTGTTTGGTTGGGTTTTGTTTTTGAGAAT

          2590      2600      2610      2620      2630      2640
TTTTTGCCCGTCCCTTTGGTTTCAAAAGTTCACGTCTTGGTGCCTTTTGTGTGACACG

          2650      2660      2670      2680      2690      2700
CCTTCCGATGGCTTGACAATGCGCAGATGTGAGGGACACGCTCACCTTAGCCTTAAGGGGG

          2710      2720      2730      2740      2750      2760
TAGGAGTGATGTGTTGGGGGAGGCTTGAGAGCAAAACGAGGAAGAGAGGGCTGAGCTGAGC

          2770      2780      2790      2800      2810      2820
TTTCGGTCTCCAGAATGTAAGAAGAAAAAATTTAAACAAAAATCTGAACTCTCAAAAGTC
```

FIGURE 1.7

```
      2830      2840      2850      2860      2870      2880
TATTTTTCTAAAACTGAAAAATGTAAATTTATACATCTATTCAGGAGTTGGGAGTGTTGTGGT

      2890      2900      2910      2920      2930      2940
TACCTACTGAGTAGGCTGCAGTTTTTGTATGTTATGAACATGAAGTTCATTATTTGTGG

      2950      2960      2970      2980      2990      3000
TTTTATTTTACTTGTACTTGTGTTGTTGCTTAAACAAGTAACCTGTTTGGCTTATAAACA

      3010      3020      3030      3040      3050      3060
CATTGAATGCGCTCTATTGCCCATGGGGATATGTGGTGTGTATCCTTCAGAAAATTAAAA

      3070      3080
GGAAAATAAAAAAAAAAAAAAAA
```

FIGURE 1.8

SUBSTITUTE SHEET

FIGURE 1B

```
                                                              R    F
"Zinc finger"        T  G  E  K  P  Y  X  C  X  X  X  X  X  F  X  X  X  X  X  X  L  X  X  H  X  X  X  H
consensus sequence


Murine Egr-1         P  H  E  R  P  Y  A  C  P  V  E  S  C  D  R  R  F  S  R  S  D  E  L  T  R  H  I  R  I  H
                     T  G  Q  K  P  F  Q  C  -  R  I  C  -  M  R  N  F  S  R  S  D  H  L  T  T  H  I  R  T  H
                     T  G  E  K  P  F  A  C  -  D  I  C  -  G  R  K  F  A  R  S  D  E  R  K  R  H  T  K  I  H


Drosophila Kruppel   S  R  D  K  S  F  T  C  K  I  -  -  C  S  R  S  F  G  Y  K  H  V  L  Q  N  H  E  R  T  H
                     T  G  E  K  P  F  F  C  P  E  -  -  C  D  K  R  F  T  R  D  H  H  L  K  T  H  M  R  L  H
                     T  G  E  K  P  Y  H  C  S  H  -  -  C  D  R  Q  F  V  Q  V  A  N  L  R  R  H  L  R  V  H
                     T  G  E  R  P  Y  T  C  E  I  -  -  C  D  G  K  F  S  D  S  N  Q  L  K  S  H  M  L  V  H


Xenopus TFIIIA       T  G  E  K  P  F  P  C  K  E  E  G  C  E  K  G  F  T  S  L  H  H  L  T  R  H  S  L  T  H
(finger 2)
```
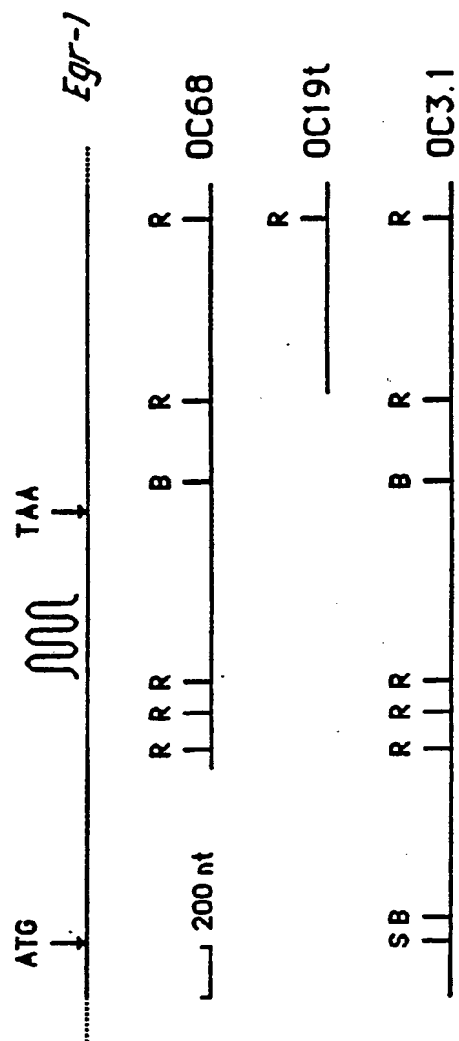
FIGURE 2

SUBSTITUTE SHEET

```
                10        20        30        40        50        60
TTTTTTTTTGGTGTGTGTGGTGGTTGTGTTTTTTAAGTGTGGAGGGCAAAAGGAGATACCA

                70        80        90       100       110       120
TCCCAGGCTCAGTCCAACCCCTCTCCAAAACNGTGTCTTTTCTGACACTCCAGGTAGCGA

               130       140       150       160       170       180
GGGAGTTGGTGTCTCCAGGTTGTGCCGAGGAGCAAATGATGACCGCCAAGGCCGTAGACAAA
                                        MetMetThrAlaLysAlaValAspLys

               190       200       210       220       230       240
ATCCCAGTAACTCTCAGTGGTTTGTGCACCAGCTGTCTGACAACATCTACCCGGTGGAG
IleProValThrLeuSerGlyPheValHisGlnLeuSerAspAsnIleTyrProValGlu

               250       260       270       280       290       300
GACCTCGCGCCCAGCTCGGTGACCATCTTTCCCAATGCCGAACTGGAGGCCCCTTTGAC
AspLeuAlaAlaThrSerValThrIlePheProAsnAlaGluLeuGlyGlyProPheAsp

               310       320       330       340       350       360
CAGATGAACGGAGTGGCCGGAGATGGCATGATCAACATTGACTGGAGAGAAGAGAGG
GlnMetAsnGlyValAlaAlaGlyAspGlyMetIleAsnIleAspMetThrGlyGluLysArg
```

FIGURE 3.1

```
370        380        390        400        410        420
TCGTTGGATCTCCCATATCCCAGCAGCTTTGCTCCCGTCTCTGCACCTAGAAACCAGACC
SerLeuAspLeuProTyrProSerPheAlaProValSerAlaProArgAsnGlnThr

430        440        450        460        470        480
TTCACTTACATGGGCAAGTTCTCCATTGACCCACAGTACCCTGGTGCCAGCTGCTACCCA
PheThrTyrMetGlyLysPheSerIleAspProGlnTyrProGlyAlaSerCysTyrPro

490        500        510        520        530        540
GAAGGCCATAATCAATATTGTGAGTGCAGGCATCTTGCAAGGGGTCACTTCCCAGCTTCA
GluGlyIleIleAsnIleValSerAlaGlyIleLeuGlnGlyValThrSerProAlaSer

550        560        570        580        590        600
ACCACAGCCTCATCCAGCGTCACCTCTGCCTCCCCAACCACTGGCCACAGGACCCCTG
ThrThrAlaSerSerValThrSerAlaSerProAsnProLeuAlaThrGlyProLeu

610        620        630        640        650        660
GGTGTGTGCACCATGTCCCAGACCCTGGACCCACCTGTACTCTCCGCCACCG
GlyValCysThrMetSerGlnThrGlnProAspLeuAspHisLeuTyrSerProPro

670        680        690        700        710        720
CCTCCTCCTTATTCTGGCTGTGCAGGAGACCTCTACCAGGACCCTTCTGGTTCCTG
ProProProTyrSerGlyCysAlaGlyAspLeuTyrGlnAspProSerAlaPheLeu
```

FIGURE 3.2

```
                730        740        750        760        770        780
TCAGCAGCCACCACCTCCACCTCTTCCTCCTCTGGCCTACCCACCCACCTCCTTCCTATCCA
SerAlaAlaThrThrSerThrSerSerLeuAlaTyrProProProSerTyrPro

                790        800        810        820        830        840
TCCCCCAAGCCAGCCACGGACCCAGGTCTCTTCCCAATGATCCCAGACTATCCTGGATTC
SerProLysProAlaThrAspProGlyLeuPheProMetIleProAspTyrProGlyPhe

                850        860        870        880        890        900
TTTCCATCTCAGTGCCAGAGAGACCTACAGCTGGTACAGCTGGCCCAGACCGTAAGCCCTTT
PheProSerGlnCysGlnArgAspLeuHisGlyThrAlaGlyProAspArgLysProPhe

                910        920        930        940        950        960
CCCTGCCCACTGGACACCCTGCGGGTGCCCCCTCCACTCCACTCTCTACAATCCGT
ProCysProLeuAspThrLeuArgValProProProLeuThrProLeuSerThrIleArg

                970        980        990        1000       1010       1020
AACTTTACCCTGGGGGCCCCAGTGCTGGGATGACCGGACCAGGGGCCAGTGGAGGCAGC
AsnPheThrLeuGlyGlyProSerAlaGlyMetThrGlyProGlyAlaSerGlyGlySer

                1030       1040       1050       1060       1070       1080
GAGGGACCCCGGCTGCCTGGTAGCAGCTCAGCAGCAGCAGCCGCCGCCGCCGCCGCCGCC
GluGlyProArgLeuProGlySerSerSerAlaAlaAlaAlaAlaAlaAlaAlaAlaAla
```

FIGURE 3.3

```
          1090      1100      1110      1120      1130      1140
TATAACCCACCACCACCTGCCACTGCGGCGCCACTGCGGCCCATTCTGAGGCCTCGCAAGTACCCCAACAGA
TyrAsnProHisHisLeuProLeuArgProIleLeuArgProLeuArgProArgLysTyrProAsnArg

          1150      1160      1170      1180      1190      1200
CCCAGCAAGACGCCGGTGCACGAGAGCCCTACCCGTGCCCAGCAGAAGGCTGCGACCGG
ProSerLysThrProValHisGluArgProTyrProCysProAlaGluGlyCysAspArg

          1210      1220      1230      1240      1250      1260
CGGTTCTCCCGCTCTGACGAGCTGACACGGCACACATCCGAATCCACACTGGGCATAAAGCCC
ArgPheSerArgSerAspGluLeuThrArgHisIleArgIleHisThrGlyHisLysPro

          1270      1280      1290      1300      1310      1320
TTCCAGTGTGCGGATCTGCATGGCCAACTTCAGCCGCCAGTGACCACCTCACCACCATATC
PheGlnCysArgIleCysMetArgAsnPheSerArgSerAspHisLeuThrThrHisIle

          1330      1340      1350      1360      1370      1380
CGCACCCACACCGGTGAGAAGCCCTTCGCCTGTGACTACTGTGGCCGAAAGTTTGCCCGG
ArgThrHisThrGlyGluLysProPheAlaCysAspTyrCysGlyArgLysPheAlaArg

          1390      1400      1410      1420      1430      1440
AGTGATGAGAGGAAGCGCCACCAAGATCCACCACCTGAGACAGAAAGAGCGGAAAAGCAGT
SerAspGluArgLysArgHisThrLysIleHisLeuArgGlnLysGluArgLysSerSer
```

SUBSTITUTE SHEET

FIGURE 3.4

```
      1450      1460      1470      1480      1490      1500
GCCCCTCTGCATCGGTGCCAGCCCCCTCTACAGCCTCCTGCTCTGGGGCGGTGCCAGGCC
AlaProSerAlaSerValProAlaProSerThrAlaSerCysSerGlyGlyValGlnAlaAla

      1510      1520      1530      1540      1550      1560
TGGGGGTACCCTGTGCAGCAGTAAACAGCAGCAGTCTTGGCGGAGGGCCGCTCGCCCCTTG
TrpGlyTyrProValGlnGlnEnd

      1570      1580      1590      1600      1610      1620
CTCCTCTCGGACCCGGACACCTTGAGATGAGACTCAGGCTGATACACCAGCTCCCAAAGG

      1630      1640      1650      1660      1670      1680
TCCCGGAGGCCCTTTGTCCACTGGAGCTGCACAAACAACTACCACCCTTTCCTGTCCC

      1690      1700      1710      1720      1730      1740
TCTCTCCCCTTGTTGGGCAAAGGGCTTTGGTGGAGCTAGCACTGCCCCCTTCCACCTAG

      1750      1760      1770      1780      1790      1800
AAGCAGGTTCTTCCTAAAACTTAGCCCCATTCTAGTCTCTTAGGTGAGTTGACTATCAA
```

FIGURE 3.5

SUBSTITUTE SHEET

```
1810      1820      1830      1840      1850      1860
CCCAAGGCAAAGGGGGAGGCTCAGAAGGAGGGTGGTGTGTGGGGATCCCCTGGCCAAGAGGGCT

1870      1880      1890      1900      1910      1920
GAGGTCTGACCCTGCTTAAAGGGTTGTTTGTTTGACTAGGTTTTGCTACCCCACTCCCCTTA

1930      1940      1950      1960      1970      1980
TTTTGACCCCATCACAGGTTTTTGACCCCTGGATGTCAGAGTTGATCTAAGACGTTTCTAC

1990      2000      2010      2020      2030      2040
AATAGGTTGGGAGATGCTGATCCCTTCAAGTGGGACAGCAAAAAGACAAGCAAAACTGA

2050      2060      2070      2080      2090      2100
TGTGCACTTTATGGCTTGGGACTGATTTGGGGGACATTGTACAGTGAGTGAAGTATAGCC

2110      2120      2130      2140      2150      2160
TTTATGCCACACTCTGTGGCCCTAAAATGGTGAATCAGAGCATATCTCTAGTTGTCTCAACC
```

FIGURE 3.6

```
      2170      2180      2190      2200      2210      2220
CTTGAAGCAATATGTATTATTATACTCAGAGAACAGAGAAGTGCAATGTGATGGGAGGAACGT

      2230      2240      2250      2260      2270      2280
AGCAATATCTGCTCCTTTCGAGTTGTTTGAGAAATGTAGGCTATTTTTTCAGTGTATAT

      2290      2300      2310      2320      2330      2340
CCACTCAGATTTTTGTGTATTTTTGATGTACCCACACTGTTCTCTAAATTCTGAATCTTTG

      2350      2360      2370      2380      2390      2400
GGAAAAATGTAAAGCATTTATGATCTCAGAGGTTAACTTATTTAAGGGGGATGTACATA

      2410      2420      2430      2440      2450      2460
TTCTCTGAAACTAGGATGCATGCAATTGTGTTGGAAGTGTCCTTGGTCGCCTTGTGTGAT

      2470      2480      2490      2500      2510      2520
GTAGACAAATGTTACAAGGCTGCCATGTAAATGGGTTGCCTTATTATGGAGAAAAAATCA
```

FIGURE 3.7

```
2530      2540      2550      2560      2570      2580
CTCCCTGAGTTTAGTATGGCTGTATATTTATGCCTATTAATAATTCAAATTTTTTTAG

2590      2600      2610      2620      2630      2640
AGTATATTTTTGTATGCTTTGTTTGTGACTTAAAAGTGTTACCTTTGTAGTCAAATTTC

2650      2660      2670      2680      2690      2700
AGATAAGAATGTACATAAATGTTACCGGAGCTGANNNTTGTTTGGTCATTAGCTCTTAATA

2710      2720      2730      2740      2750      2760
GTTGTGAAAAATAAATCTATTCTAACGCAAAACCACTAACTGAAGTTCAGATATAAATGG

2770      2780      2790      2800      2810
ATGGTTTGTGACTATAGTGTAAATAAATACTTTTCAACAAAAAAAAAAAAAAAA
```

FIGURE 3.8

```
-935                                                              -876
ACGGAGGGGAA TAGCCTTTCG ATTCTGGGTG GTGCATTGGA AGCCCCAGGC TCTAAAACCC

-875                                                              -816
CCAACCTACT GACTGGTGGC CGAGTATGCA CCCGACTGCT AGCTAGGCAG TGTCCCAAGA

-815                                                              -756
ACCAGTAGCC AAATGTCTTG GCCTCAGTTT TCCCGGTGAC ACCTGGAAAG TGACCCTGCC

-755                                                              -696
ATTAGTAGAG GCTCAGGTCA GGGCCCCGCC TCTCCTGGGC GGCCTCTGCC CTAGCCCGCC

-695                                                              -636
CTGCCGCTCC TCCTCTCCGC AGGCTCGCTC CCACGGTCCC CGAGGTGGGC GGGTGAGCCC

-635                                                              -576
AGGATGACGG CTGTAGAACC CCGGCCTGAC TCGCCCTCGC CCCCGCGCCG GGCCTGGGCT

-575                                                              -516
TCCCTAGCCC AGCTCGCACC CGGGGGCCGT CGGAGCCGCC GCGCGCCCAG CTCTACGCGC

-515                                                              -456
CTGGCCCTCC CCACGCGGGC GTCCCCGACT CCCGCGCGCG CTCAGGCTCC CAGTTGGGAA

-455                                                              -396
CCAAGGAGGG GGAGGATGGG GGGGGGGGTG TGCGGCGACC CGGAAACGCC ATATAAGGAG

-395                                                              -336
CAGGAAGGAT CCCCCGCCGG AACAGGACCTT ATTTGGGCAG CGCCCTTATAT GGAGTGGCCC
```

FIGURE 4.1

-335
ÀATAATGGCCC TGCCGCTTCC GGCTCTGGGA GGAGGGGCGA GCGGGGGTTG GGGCGGGGGGC
-276

-275
AAGCTGGGAA CTCCAGGCGC CTGGCCCGGG AGGCCACTGC TGCTGTTCCA ATACTAGGCT
-216
SmaI

-215
TTCCAGGAGC CTGAGCGCTC GCGATGCCGG AGCGGGTCGC AGGGTGGAGG TGCCCACCAC
-156

-155
TCTTGGATGG GAGGGCTTCA CGTCACTCCG GGTCCTCCCG GCCGGTCCTT CCATATTAGG
-96

-95
GCTTCCTGCT TCCCATATAT GGCCATGTAC GTCACGGCGG AGGCGGGCCC GTGCTGTTCC
-36

-35
AGACCCTTGA AATAGAGGCC GATTCGGGGA GTCGCGAGAG ATCCCAGCGC GCAGAACTTG
+25
                        +1
         29-mer

+26
GGGAGCGCC GCCGCGATTC GCCGCGGCCG CCAGCTTCCG CCGGCGCAAG ATCGGCCCCT
+85

+86
GCCCCAGCCT CCGCGGCAGC CCTGCGTCCA CCACGGGGCG CGGCTACCGC CAGCCTGGGG
+145

+146
GCCCACCTAC ACTCCCCGCA GTGTGCCCCT GCACCCCGCA TGTAACCCGG CCAACCCCCG
+205

+206
GCGAGTGTGC CCTCAGTAGC TTCGGCCCCG GGCTGCGCCC ACCACCCAAC ATCAGTTCTC
+265
SmaI

FIGURE 4.2
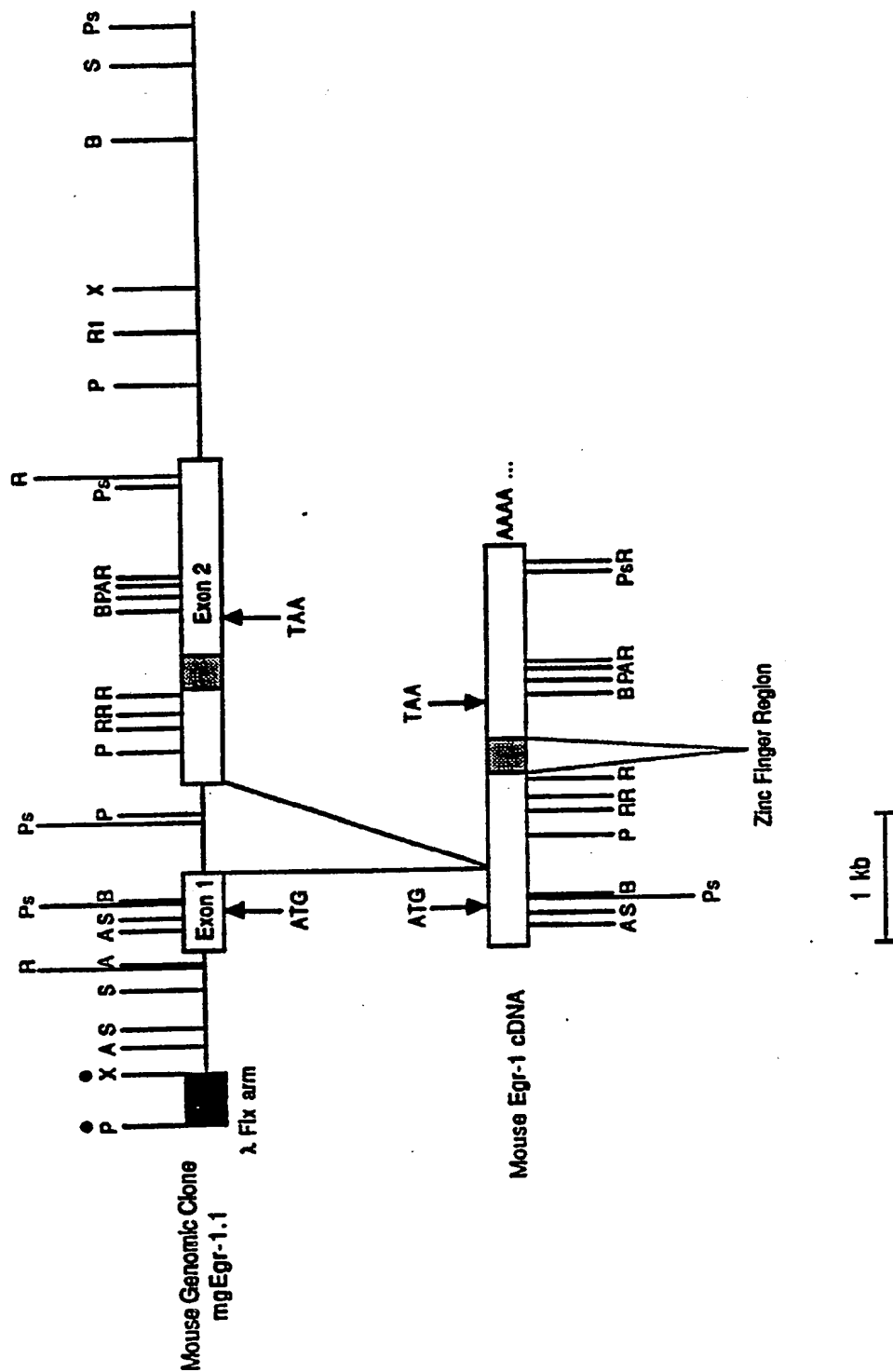
FIGURE 5

INTERNATIONAL SEARCH REPORT

International Application No. **PCT/US89/01473**

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) 6

According to International Patent Classification (IPC) or to both National Classification and IPC

CO7H 15/12, C12P 21/00, 19/34; C12N 15/00, 7/00, 1/20;
C12Q 1/68; CO7K 13/00

II. FIELDS SEARCHED

Minimum Documentation Searched 7

Classification System	Classification Symbols	
US	536/27; 435/68, 91, 252.3, 6;530/350, 387	172.1, 172.3, 320,

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched 8

CAS database (1967-1989) keywords: finger/protein/motif
early growth regulatory/gene/expression/sequence

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
X Y	Chavrier, P. et al (January 1988) EMBO Journal, Volume 7: 29-35;" A gene encoding a protein with zinc fingers is activated during G₀/G₁, transition in cultured cells. See entire document.	1-5,8-10 6,7,11,12 14-16
X Y	Chowdhury, K. et al (March 1988) Cell, Volume 48: 771-778; "A Multigene family encoding several finger structures is present and differentially active in mammalian genomes". See entire document.	1-5,8-10 6,7,11, 12,14-16
X Y	Lau, L. et al (March 1987) PNAS, Volume 84: 1182-1186; "Expression of a set of growth-related immediate early genes in BALB/c 3T3 cells: coordinate regulation with c-fos or c-myc. See entire document	1-5,7-11 6,12, 14-16

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral-disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search

27 June 1989

Date of Mailing of this International Search Report

24 AUG 1989

International Searching Authority

ISA US

Signature of Authorized Officer

Anne Brown

Form PCT/ISA/210 (second sheet) (Rev.11-87)

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

, P,Y	Pannuti, A. et al (May 1988) Nucleic Acids Res., Volume 16: 4227-4237 "Isolation of CDNAs encoding finger proteins and measurment of the corresponding mRNA levels during myeloid terminal differentiation" See entire document.	1-12, 14-16
P,Y	Almendrol et al (May 1988) Mol Cell Biol, Vol. 8:2140-2148 "Complexity of the Early Genetic Response to Growth Factors in Mouse Fibroblasts". See entire document.	1-12, 14-16

V.☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1.☐ Claim numbers   . because they relate to subject matter [12] not required to be searched by this Authority, namely:

2.☐ Claim numbers _..__ . because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [13], specifically:

3.☐ Claim numbers_____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI.☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

See attached sheet

1.☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2.☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3.☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers: 1-12, 14-16

4.☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (supplemental sheet (2) (Rev. 11-87)

ATTACHMENT TO PCT/ISA/210
VI.   OBSERVATION WHERE UNITY OF INVENTION IS LACKING

I.   Claims 1-12, and 14-16 are drawn to EGR DNA
sequence, vector or cell containing it, transformed
host, method of using vector or cell and a polypeptide.

II.   Claim 13 is drawn to cell-free method of pre-
paring an early growth regulatory protein.

III.   Claims 17-19 are drawn to native EGR protein.

IV.   Claims 20-23 are drawn to synthetic peptide
fragment antigenically related to and containing homo-
logy to native EGR protein.

V.   Claims 24-29 are drawn to antibodies to EGR,
anti-bodies to region EGR, method of using antibody.

VI.   Claim 30 is drawn to method of detecting mRNA.

VII.   Claim 31 is drawn to method of detecting DNA.

VIII.   Claim 32 is drawn to method of diagnosis.

The claims do not embrace one single general inven-
tive concept as defined in Rule 13.

Groups I, III, IV and V contain claims directed to
distinct chemical entities. These are a gene and a pro-
tein coded for by that gene, a native EGR protein, a
synthetic peptide fragment, antibodies which bind to EGR
proteins. Groups II, VI, VII and VIII are directed to
alternative uses for the nucleotide defined in Group I.